A8

7/5/1 (Item 1 from file: 347)
DIALOG(R)File 347:JAPIO

04233540 **Image available**
DOCUMENT DATA BASE DEVICE

PUB. NO.: 05-225240 **JP 5225240** A]
PUBLISHED: September 03, 1993 (19930903)
INVENTOR(s): OKUMURA HIROSHI
APPLICANT(s): FUJI XEROX CO LTD [359761] (A Japanese Company or Corporation), JP (Japan)
APPL. NO.: 04-234911 [JP 92234911]
FILED: September 02, 1992 (19920902)
INTL CLASS: [5] G06F-015/40; G06F-015/20
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)
JOURNAL: Section: P, Section No. 1659, Vol. 17, No. 677, Pg. 86, December 13, 1993 (19931213)

ABSTRACT

PURPOSE: To extract partial document contents from a structured document stored in a document data base.

CONSTITUTION: A document retrieval means 12 reports information which designates a set of documents out of information, which are reported from a retrieval expression analysis means 11, to a document set retrieval means 13 and reports information related to extraction of the document structure out of these information to a document structure extracting means 14. The document set retrieval means 13 retrieves a set of documents in a document storage part 2 based on reported information. The document structure extracting means 14 extracts the document structure of each element of the set of documents retrieved by the document set retrieval means 13. A document allocating means 15 allocates unallocated documents extracted and reconstituted by the document structure extracting means 14.

1/1件

JAPANESE PATENT OFFICE

公開特許公報フロントページ

(11)公開番号: 特開平05-225240
(43)公開日: 1993年09月03日

(51) Int. Cl. 5 G06F 15/40 500
15/20 550

(21)出願番号: 特願平04-234911
(22)出願日: 1992年09月02日

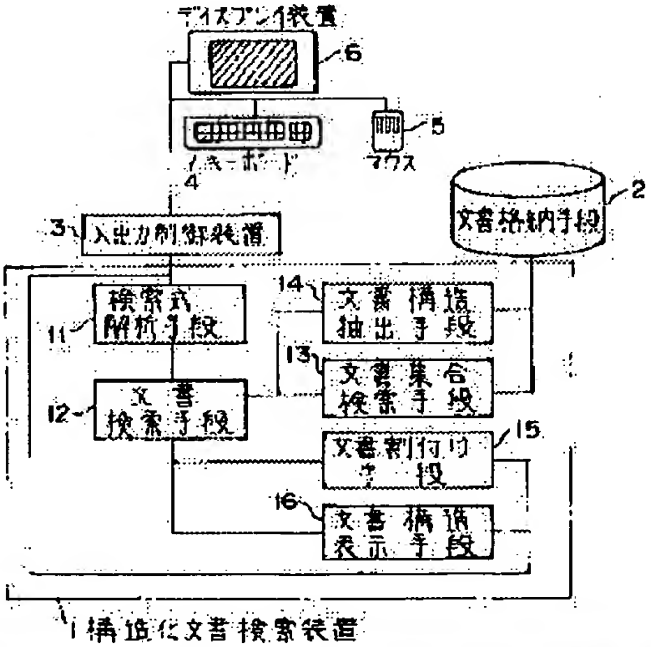
(71)出願人: 富士ゼロックス株式会社
(72)発明者: 奥村洋

(30)優先権
優先権主張番号: 1991246220 優先日: 1991年09月25日 優先権主張国: JP

(54) 文書データベース装置

(57)要約:

【目的】文書データベースに蓄積されている構造化文書から一部分の文書内容を抽出する。
【構成】文書検索手段１２は検索式解析手段１１から通知された情報のうち、文書集合を指定する情報を文書集合検索手段１３に通知すると共に、文書構造の抽出に関する情報を文書構造抽出手段１４に通知する。文書集合検索手段１３では、通知された情報に基づいて、文書格納手段２内の文書集合を検索する。また文書構造抽出手段１４では、通知された情報に基づいて、文書集合検索手段１３により検索された文書集合の各要素の文書構造を抽出する。更に文書割付け手段１５は、文書構造抽出手段１４によって抽出され再構成された割付けされていない文書を、割付けする。



リーガルステータス

【審査請求日】 1995年09月07日
【拒絶査定発送日】 1998年09月29日
【最終処分種別】
【最終処分日】
【特許番号】
【登録日】
【拒絶査定不服審判番号】
【拒絶査定不服審判請求日】
【本権利消滅日】

1

【特許請求の範囲】
【請求項1】構造化文書の文書構造の抽出に関する情報を指定する指定手段と、
前記指定手段により指定された前記情報に基づいて、検索対象の構造化文書の文書構造を抽出する文書構造抽出手段とを具えたことを特徴とする文書データベース装置。
【請求項2】複数の構造化文書を格納する文書格納手段と、
文書集合を指定する情報と文書構造の抽出に関する情報とを指定する指定手段と、
前記文書格納手段から、前記文書集合を指定する情報に適合する文書集合を検索する文書集合検索手段と、
前記文書構造の抽出に関する情報に基づいて、前記文書集合検索手段により検索された文書集合の各要素の文書構造を抽出する文書構造抽出手段とを具えたことを特徴とする文書データベース装置。
【請求項3】前記文書構造抽出手段により抽出された文書構造を割付ける文書割付け手段を更に具えたことを特徴とする請求項1又は2記載の文書データベース装置。
【請求項4】前記文書構造抽出手段により抽出された文書構造を所定の表示形式に従って表示する文書構造表示手段を更に具えたことを特徴とする請求項1又は2記載の文書データベース装置。
【請求項5】文書構造をそれぞれ異なった表示形式で表示する複数の文書構造表示手段と、当該各文書構造表示手段を1つ以上選択する表示選択手段とを更に具えたことを特徴とする請求項1又は2記載の文書データベース装置。
【請求項6】前記文書割付け手段は、前記文書構造抽出手段により抽出された文書構造が属していた構造化文書に対応する文書割付けテンプレートに従って、前記抽出された文書構造を割付けることを特徴とする請求項3記載の文書データベース装置。
【請求項7】複数の文書割付けテンプレートを格納するテンプレート格納手段と、前記各文書割付けテンプレート中の所望の文書割付けテンプレートを指定するテンプレート指定手段とを更に具え、前記文書割付け手段が、前記テンプレート指定手段により指定された文書割付けテンプレートに基づいて、前記抽出された文書構造を割付けることを特徴とする請求項3記載の文書データベース装置。
【請求項8】前記テンプレート格納手段に格納されている文書割付けテンプレートを編集するテンプレート編集手段を更に具えたことを特徴とする請求項6記載の文書データベース装置。
【発明の詳細な説明】
【0001】
【産業上の利用分野】この発明は、コンピュータを利用した文書処理、特に既存文書の再利用による文書処理を

2

可能にする文書データベース装置に関する。
【0002】
【従来の技術】従来、計算機を利用した文書データベースは、特にオフィスを中心として、文書を蓄積・再利用することを目的として多くのユーザによって利用されている。文書を蓄積・再利用するための装置としては、文書ファイリング装置があったが、文書ファイリング装置では文書の格納場所が分からない場合は所望の文書を得る事が出来ず、文書データベースとして利用することができなかった。
【0003】このような問題を解決すべく、従来の文書データベース装置は文書ファイリング装置と文書検索装置とから構成されていた。このような文書データベース装置によって処理される文書には属性が登録されている。その属性は、文書ファイリング装置に文書が蓄積される前にその文書に登録されるようになっている。また属性の登録に際しては、システムによって自動的になされても良いし、ユーザによって手動でなされても構わない。
【0004】ここで、属性とは、例えばユーザが与えるキーワード、システムによって自動的に登録されるキーワード、文書自体が持っている情報である作者や作成年月日などである。このような属性はインデックスとして登録される。
【0005】そして、文書検索装置は、検索時には、上述した属性をキーとしてインデックスを用いて、文書ファイリング装置内を検索することにより、結果として文書の存在場所、名前を得るとともに、その結果に基づいて文書ファイリング装置から所望の文書全体を取り出して、それを検索結果としていた。
【0006】しかしながら、上記従来の文書データベース装置では文書全体を取り出すことしかできないという問題があった。
【0007】そのため、例えば文書の一定の部分のみを必要とするユーザやアプリケーションは、一旦、文書をリモートの文書ファイリング装置からローカル的な領域（例えば記憶装置）に取り出した後、その文書全体を解析し、その後、解析結果に基づいて必要な部分（文書内容）のみを取り出さなければならない。このため、転送されるデータ量が必要なデータ量に比べて非常に多くなるので転送効率が悪くなってしまうとともに、文書全体が得られてしまうので得られた文書を再編成する必要が生じてしまい、ユーザにとって非常に大きな負担となっていた。
【0008】このような欠点を解決するために、関係データベースに代表される2次元の表を扱えるシステムを利用したり、文書内にタグを入れたりすることによって文書の内容を取り出したりする工夫もなされた。このようなシステムでは、文書を2次元の表として扱い、その表を関係データベースに埋め込んだり、あるいはタグを

文書中に埋め込んだりすることによって一部分を取り出すという機能を実現していた。表は、例えば（タイトル：" 文書データベース装置" ）というように、（キー：値）という「キー」と「値」の組みからなる組み合わせの列として表現される。

【０００９】また、近年においては、上記のような構造を持たない文書ではなく、論理的な構造を持った構造化文書が扱われるようになった。文書が、例えば国際規格ＯＤＡ（Open Document Architecture；ISO 8613）に準拠している場合、図１３に示すような文書は、図１４に示すような、フレーム内にブロックが配置されている入れ子構造の割付け構造を持ち、かつ、図１５に示す様な内部表現（文書構造）を持っている。図１５に示す文書構造において、図中矢印Ａで示す点線より上部が文書の論理構造であり、また図中矢印Ｂで示す点線より下部が文書の割付け構造であり、更に図中矢印Ａで示す点線と図中矢印Ｂで示す点線との間に位置している部分が文書の内容（内容部）である。また文書の割付け構造は、図１６に示割付けテンプレート（ＯＤＡでは共通割付け構造という）を用いて、文書の論理構造から生成される。これを割付けという。

【００１０】なお内部構造は、図１５に示すように、更に細かい文書部品から構成されており、各々の文書部品は親の文書部品、子の文書部品、孫の文書部品という様な親子関係の木構造を有している。例えば図１５に示される、論理構造を構成する各構成要素および割付け構造を構成する各構成要素はそれぞれ文書部品である。

【００１１】更に図１５に示すような文書構造は実際には、図１７に示すように、文書内容を保持（実際にはポインタなどによって指し示している）している論理構造と、図１８に示すように、文書内容を保持（実際にはポインタなどによって指し示している）している割付け構造とで表現される。

【００１２】しかしながら、従来の文書データベース装置では、文書単位でしか文書を取り扱えなかったので、上記ＯＤＡに準拠している文書の如く、構造化文書の文書構造を取り扱うことができなかった。

【００１３】そのため次のような問題が生じていた。

【００１４】（１）文書が階層のない２次元の表として扱われていたため、文書の階層的な論理構造を扱うことができなかった。例えば、ある文書の論理構造がタイトル、著者名、段落から構成され、その段落が段落タイトル、段落内容から構成されるものとする。そして、このような文書の段落の段落タイトルを取り出すという検索を行う場合、従来の文書データベース装置では、文書全体を取り出すか、あるいは文書中の段落全体を取り出さなければならなかった。そのため、文書を取り出した後の文書編集に手間がかかったり、文書の転送のデータ量が多く転送効率が悪いという問題があった。

【００１５】（２）複数の文書の検索結果として一つの文書を得る場合、従来の文書データベース装置では、複数の文書から著者のみを取り出して一つの文書とする、といったように検索結果を「表」として表現することしかできず、「表」では表現できない階層構造を持った文書を得ることはできなかった。このため、文書全体を取り出した後、その文書を編集して一つの文書を作り出すという作業が発生してしまうこととなり、その文書を取り出した後に文書編集を行わなければならず、文書編集に手間がかかったり、ユーザに大きな負担を強いることになっていた。また文書の転送のデータ量が多く転送効率が悪いという問題もあった。

【００１６】

【発明が解決しようとする課題】このように上記従来の文書データベース装置では、構造化文書の文書構造を取り扱うことができず、文書単位でしか文書を取り扱うことができなかったので、文書全体を取り出してから、その文書を編集して一つの所望の文書を作り出すという作業を行わなければならなかった。このため、取り出した文書の内容編集及び割付けに手間がかかることとなり、ユーザに大きな負担がかかっていた。

【００１７】また文書中の不必要な情報も同時に伝送されるため、データ伝送量が多く、伝送効率が悪いという欠点があった。

【００１８】そこで、本発明は、文書データベースに蓄積されている構造化文書から一部分の文書内容を抽出することができると共に、該抽出した内容の割付け処理を行うことができ、かつ、伝送データ量、伝送時間、文書編集時間及び文書割付け時間を軽減することのできる文書データベース装置を提供することを目的とする。

【００１９】

【課題を解決するための手段】上記目的を達成するため、第１の発明の文書データベース装置は、構造化文書の文書構造の抽出に関する情報を指定する指定手段と、該指定手段により指定された前記情報に基づいて、検索対象の構造化文書の文書構造を抽出する文書構造抽出手段とを具えている。

【００２０】第２の発明の文書データベース装置は、複数の構造化文書を格納する文書格納手段と、文書集合を指定する情報と文書構造の抽出に関する情報とを指定する指定手段と、前記文書格納手段から、前記文書集合を指定する情報に適合する文書集合を検索する文書集合検索手段と、前記文書構造の抽出に関する情報に基づいて、前記文書集合検索手段により検索された文書集合の各要素の文書構造を抽出する文書構造抽出手段とを具えている。

【００２１】第３の発明の文書データベース装置は、第１の発明又は第２の発明において、前記文書構造抽出手段により抽出された文書構造を割付ける文書割付け手段を更に具えたことを特徴とする。

【００２２】第４の発明の文書データベース装置は、第

1の発明又は第2の発明において、前記文書構造抽出手段により抽出された文書構造を所定の表示形式に従って表示する文書構造表示手段を更に具えている。

【0023】第5の発明の文書データベース装置は、第1の発明又は第2の発明において、文書構造をそれぞれ異なった表示形式で表示する複数の文書構造表示手段と、当該各文書構造表示手段を1つ以上選択する表示選択手段とを更に具えている。

【0024】第6の発明の文書データベース装置は、第3の発明において、前記文書構造抽出手段により抽出された文書構造が属していた構造化文書に対応する文書割付けテンプレートに従って、前記抽出された文書構造を割付けることを特徴としている。

【0025】第7の発明の文書データベース装置は、第3の発明において、複数の文書割付けテンプレートを格納するテンプレート格納手段と、前記各文書割付けテンプレート中の所望の文書割付けテンプレートを指定するテンプレート指定手段とを更に具え、前記文書割付け手段が、前記テンプレート指定手段により指定された文書割付けテンプレートに基づいて、前記抽出された文書構造を割付けることを特徴とする。

【0026】第8の発明の文書データベース装置は、第6の発明において、前記テンプレート格納手段に格納されている文書割付けテンプレートを編集するテンプレート編集手段を更に具えている。

【0027】

【作用】第1の発明によれば、検索対象の構造化文書から、構造化文書の抽出に関する情報に基づいた文書構造が抽出されるので、構造化文書の一部分のみを抽出することができる。

【0028】第2の発明によれば、複数の構造化文書中から、文書集合を指定する情報に適合する文書集合が検索され、更に、その文書集合の各要素（検索された各構造化文書）から、構造化文書の抽出に関する情報に基づいた文書構造が抽出されるので、複数の構造化文書からそれぞれ、同一の文書構造を有する一部分の内容のみを抽出することができる。

【0029】第3の発明によれば、文書割付け手段によって、文書構造抽出手段により抽出された文書構造の内容を割付けるようにしたので、割付け済みの文書を得ることができる。

【0030】第4の発明によれば、文書構造表示手段によって、文書構造抽出手段により抽出された文書構造の内容を表示するようにしたので、抽出された文書構造の内容を視覚的に認識することができる。

【0031】第5の発明によれば、表示選択手段によって、複数の文書構造表示手段中から単数又は複数の文書構造表示手段を選択し、この選択された文書構造表示手によって、文書構造抽出手段により抽出された文書構造の内容を表示するようにしたので、所望の表示形式で、

抽出された文書構造の内容を視覚的に認識することができる。

【0032】第6の発明によれば、文書割付け手段によって、文書構造抽出手段により抽出された文書構造が属していた構造化文書に対応する予め設定された文書割付けテンプレートに従って、文書構造抽出手段により抽出された文書構造を割付けるようにしたので、割付け済みの文書を得ることができる。

【0033】第7の発明によれば、文書割付け手段が、テンプレート指定手段により指定された文書割付けテンプレートに従って、文書構造抽出手段により抽出された文書構造を割付けるようにしたので、割付け済みの文書を得ることができる。

【0034】第8の発明によれば、文書テンプレート編集手段は、テンプレート格納手段に格納されている文書割付けテンプレートを作成／削除／変更する。文書割付け手段は、文書テンプレート編集手段により編集された文書割付けテンプレートに従って、文書構造抽出手段により抽出された文書構造を割付けるようにしたので、所望のレイアウトに応じた割付け済みの文書を得ることができる。

【0035】

【実施例】以下、本発明の実施例を添付図面を参照して説明する。

【0036】最初に本発明の第1の実施例を図1乃至図11を参照して説明する。

【0037】図1は本発明に係る文書データベース装置の第1の実施例を機能ブロック図で示したものである。同図において、文書データベース装置は、構造化文書検索装置1、文書格納手段2、入出力制御装置3、キーボード4、マウス5、ディスプレイ装置6を備えている。

【0038】文書格納手段2は、例えば磁気ディスクを備えて構成される大容量ファイリングシステムであって、ここには図14に示したような内部構造を持った構造化文書が複数格納されている。

【0039】入出力制御装置3は、構造化文書検索装置1とキーボード4とマウス5およびディスプレイ装置6との間の入出力を制御するものである。例えばディスプレイ装置6の表示画面上で形成された文書は入出力制御装置3を通じて文書格納手段2に格納される。また構造化文書検索装置1によって文書格納手段2から読み出された構造化文書は、入出力制御装置3を通じてディスプレイ装置6へ伝送され表示画面上に表示される。

【0040】キーボード4及びマウス5は、各種データ及びコマンド等を入力するために操作されるものであり、この操作による入力に応じた表示がディスプレイ装置6の表示画面上になされる。

【0041】検索式解析手段11は、検索式（詳細は後述する）を解析するためのものであり、指示された検索式を文書構造抽出式と文書集合検索式とに分け、文書検

索手段１２に対して検索を実行すべき旨を指示する。

【００４２】文書検索手段１２は、検索を実行するためのものであり、検索すべき旨の指示に従って文書集合検索手段１３、文書構造抽出手段１４、文書割付け手段１５に所定の指示やデータ伝送を行う。

【００４３】文書集合検索手段１３は、文書格納手段２内の文書を検索するためのものであり、文書に予め与えられている属性やキーワードに基づいて、与えられた条件、つまり検索式内の文書集合検索式を満たす文書の集合を検索する。

【００４４】文書構造抽出手段１４は、構造化文書中の文書部品を検索するためのものであり、指示された部分、つまり検索式内の文書構造抽出式に基づく部分を抽出した後、必要があれば構造の変更を行う。また構造化文書に対応する文書割付けテンプレートも読み出す。

【００４５】文書割付け手段１５は、文書割付テンプレートに従って割付けを行うものであり、文書論理構造の抽出を行った結果、割付けが崩れてしまった構造化文書つまり文書論理構造に対して、文書割付テンプレートに従って新たな割付け構造を付与する。

【００４６】文書構造表示手段１６は、割付け構造を持たない構造化文書つまり文書論理構造の表示を行うためのものであり、抽出された文書論理構造をディスプレイ装置６上に表示させる。

【００４７】なお、この実施例では、文書割付け手段１５と文書構造表示手段１６とが設けられているが、これら各手段は同時に機能するようなことはなく、いずれかの手段のみが機能するように設定されている。そのために、いずれかの手段を指定でき、かつ、切り替えられるようになっている。これは、ユーザによって指定された情報が、入出力制御装置３、検索式解析手段１１を経て文書検索手段１２に入力されると、文書検索手段１２がその情報に基づいて上記各手段を切り替えるようになっている。ここで、文書割付け手段１５が機能するよう指定された場合には、抽出された文書論理構造が割付けされて表示されることとなり、一方、文書構造表示手段１６が機能するよう指定された場合は、割付け構造を持たない構造化文書つまり文書論理構造がが表示されることとなる。

【００４８】勿論、このように切り替え方式ではなく、上述した構成において、文書割付け手段１５を削除した構成や、文書構造表示手段１８を削除した構成とすることも可能である。また複数の文書構造表示手段を設けたが、１つの文書構造表示手段のみ設けるようにしても良い。

【００４９】上述した構成において、文書格納手段２内の複数の構造化文書から所望の検索結果を得るためには、検索式（検索条件）を設定しなければならないので、次にその検索式について説明する。

【００５０】図２は、検索式の一例を示したものであり、この検索式は、上述した指定手段の機能を果たすものであり、文書集合を指定する情報（文書集合検索式）と、文書構造の抽出に関する情報（文書構造抽出式）とから構成されている。

【００５１】図２に示す検索式において、「Ｆｒｏｍ」と「Ｐｒｏｊｅｃｔ」とが対になって構成されており、「Ｆｒｏｍ」２３Ａと「Ｐｒｏｊｅｃｔ」２３Ｂとが、「Ｆｒｏｍ」２４Ａと「Ｐｒｏｊｅｃｔ」２４Ｂとが、「Ｆｒｏｍ」２５Ａと「Ｐｒｏｊｅｃｔ」２５Ｂとが、それぞれ対になっている。各「Ｆｒｏｍ」は検索の範囲を指定するものであり、また各「Ｐｒｏｊｅｃｔ」は検索の範囲から抽出する部分を指定するものであり、更に「Ｃｏｌｌａｐｓｅ」２６は一段階層を浅くすることを指定している。なお「Ｆｒｏｍ」２４Ａに指定されている検索の範囲を示す情報「＊」２７は「何でも良い」ということを表している。

【００５２】「Ｐｒｏｊｅｃｔ」及び「Ｆｒｏｍ」では属性名を指定することによって範囲を指定しているが、指示される属性名は構造化文書の作成に際して、文書部品の識別が可能なように文書論理構造の文書部品に与えられるものである。なお、この属性値は必ずしも文書部品の名前でなくとも良く、作成日、作成者名等であっても構わない。

【００５３】また最も外側の「Ｆｒｏｍ」２３Ａは文書集合の指定を行う。ここには文書集合検索式を指示することができ、これは、文書集合検索手段１３に指示することのできる文書集合検索式である。この文書集合検索式は、文書の作成日、文書の種類別たとえば「特許」、「オブジェクト指向言語に関する論文」などを示す情報等、文書の属性値の指定でも構わない。また構造化文書の構造化に立ち入った指定であっても構わない。これは、構造化文書を構成している構成要素を指定することであり、指定された構成要素を有する構造化文書全てが検索対象となる。例えば構成要素としての「タイトル」として「文書データベース装置」を指定することにより、構成要素としての「タイトル」が「文書データベース装置」となっている、全ての構造化文書が検索対象となる。これにより同一技術分野の文書を検索することができる。

【００５４】図２に示される検索式によれば、最も外側の「Ｆｒｏｍ」２３Ａによって、文書集合として、文書格納手段２内のディレクトリ" ／データベース研究関連／論文" 中の、全ての文書が指示される。この文書集合の各要素（各文書）に対して「Ｐｒｏｊｅｃｔ」２３Ｂによって、タイトル、著者名、段落の抽出の指示がなされる。更に内側の「Ｆｒｏｍ」２５Ａによって段落が指示され、内側の「Ｐｒｏｊｅｃｔ」２５Ｂによって段落から更に段落タイトルを抽出することが指示される。また、「Ｃｏｌｌａｐｓｅ」２７によって段落の直下にある段落タイトルは文書論理根の直下に移動することも指

示されている。

【0055】また図2に示した検索式は、文書集合の指定と文書抽出の指定ができる検索式であればどのようなものと代替しても構わない。

【0056】上述したような検索式は、キーボード4、マウス5を操作することにより指定されディスプレイ装置6の表示画面上に表示される。こうしてディスプレイ装置6の表示画面上で検索式（検索条件）を設定すると、この検索式は、入出力制御装置3を介して検索式解析手段11に通知される。

【0057】検索式解析手段11は、例えば図2に示される検索式が通知されると、その検索式を解析して、文書集合検索式（図2に示す文書集合検索式21）と文書構造抽出式（図2に示す文書構造抽出式22）とを文書検索手段12に通知する。

【0058】なお、図2に示す文書構造抽出式22は、図3に示されるような木構造（以下解析木という）で表現することができる。従って、この実施例では、検索式解析手段11から文書検索手段12に通知される文書構造の抽出に関する情報は、文書構造抽出式22（図2参照）のような形式ではなく、検索式解析手段11によって求められる図3に示すような解析木30の形式で通知されるようになっている。勿論、文書構造抽出式22そのものを通知するようにしても良い。

【0059】さて、文書検索手段12は、文書集合検索式21及び解析木30が通知されると、文書集合検索式21を文書集合検索手段13に通知する。文書集合検索手段13では、通知された文書集合検索式21に基づいて、文書格納手段2から該当する複数の構造化文書を探し出す。勿論、場合によっては1つの文書のときも有り得る。文書集合検索手段13はその各構造化文書へのポインタの集合を文書検索手段12に指示する。そのポインタの集合の一例を図4に示す。

【0060】ここでは、文書集合検索手段13によって図4に示すように、構造化文書41A、42A、43A、44A、45Aが検索され、その各文書を指し示すポインタ41、42、43、44、45が、文書集合検索手段13から文書検索手段12に指示されるものとする。

【0061】文書検索手段12は、指示されたポインタの集合を、検索式解析手段11から既に通知されている解析木30とともに文書構造抽出手段14に指示する。すると文書構造抽出手段14は、構造化文書へのポインタの集合および解析木30が指示されると、各ポインタで指し示される構造化文書のファイル内部へポインタを張るための情報を、構造化文書毎に順次読み込むとともに、その読み込んだ情報に基づいて構造化文書に対して文書構造抽出処理を行い、その後、抽出された文書論理構造を文書検索手段12に転送する。

【0062】図5は、文書格納手段2内に格納されている構造化文書のファイル内部へ、ポインタを張るための情報の読み込みを説明するための図を示したものである。

【0063】ここでは、図4に示したポインタ41で指し示されている構造化文書41Aが図14に示した構造化文書であった場合の、上記情報の読み込みについて説明する。

【0064】この実施例では、構造化文書41Aの文書論理構造51を構成する各構成要素を参照するためのインデックスが保持されているインデックスファイル52が、文書格納手段2に予め格納されているので、文書構造抽出手段14は、ポインタ41で指し示される構造化文書41Aに対応するインデックスファイル52を自己内に読み込むことにより、ポインタを張るための情報を得ることができる。なお、この時点では構造化文書のファイルをオープンにする必要はないが、ポインタを張るときにはファイルをオープンにする必要がある。同様にして、ポインタ42～45で指し示される各構造化文書のファイル内部へポインタを張るための情報を得ることができる。

【0065】また上述した方法以外に、インデックスファイル52を設けず、構造化文書41Aのファイルをオープンして、その文書の文書論理構造51を読み込むようにしても良い。なお、このとき文書の内容（内容部）53は読み込まれない。なぜならば、一般的に、文書論理構造のデータ量に比べて文書内容のデータ量の方が非常に多いからである。

【0066】次に文書構造抽出手段14による文書構造抽出処理を説明する。

【0067】（1）最初に、読み込んだ構造化文書のファイル内部へポインタを張るための情報（この例ではインデックスファイル52）に基づいて、解析木30（図3参照）から、該当する構造化文書（この例では構造化文書41A）に対して、その解析木のラベルに従ってポインタ（リンク）を張る。そのポインタが張られた状態を図6に示す。図6から分かるように、解析木30を構成する「タイトル」、「著者名」、「段落」、「段落タイトル」の各ノードを表している各ラベルと一致する文書論理構造51の構成要素（文書部品）に対して、ポインタP1～P4が設定されている。これは、解析木30と、解析木30の要素に対応する論理構造51における構成要素とが対応付けされたポインタP1～P4が、文書構造抽出手段14内に保持されたことを意味する。なお、図6に示す状態においては、解析木30は文書構造抽出手段14内に存在しており、また文書論理構造51及び文書の内容部53は文書格納手段2内に存在している。なおここでは構造化文書41Aの割付け構造については省略してある。

【0068】（2）次に、そのリンク情報に基づいて文書論理構造をコピーする。そのコピーされた文書論理構

造の一例を図7に示す。文書構造抽出手段14は、図7に示す様に、コピーされた文書論理構造71及び文書の内容部72を、自己内に読み込む。この実施例では、文書論理構造71及び文書の内容部72のみを読み込むようにしているので、文書論理構造71に対応する割付け構造は得られない（割付け構造は存在していない）。

【0069】（3）更に、文書構造を変更するタグがついている解析木のノードがあればその指示に従って構造を変更する。

【0070】図7に示すように、解析木30には「Collapse」31のタグがついたノード段落32が存在するので、図7に示される文書論理構造71は図8のように変更される。図8から分かるように、「Collapse」31のタグによって「段落タイトル」の階層がなくなり、文書内容81が段落の「ノード」82の直下に存在している。また図8に示すような、抽出された文書論理構造は新たな構造化文書となる。

【0071】上記（1）～（3）の処理で一つの文書に対する文書構造抽出処理は終了することとなる。このとき、必ずしも全ての構造化文書から文書論理構造が抽出されるとは限らず、文書構造抽出式（解析木）に適合する文書論理構造を有する構造化文書からのみ、その文書論理構造が抽出されることとなる。従って、抽出される文書論理構造が「なし」という場合も有り得る。

【0072】ところで、文書検索手段12は構造化文書の検索処理を終了すると、転送された文書論理構造の集合が少なくとも1つ以上の文書である場合は、文書のリストを作成し、そのリストを入出力制御装置3を通じてディスプレイ装置6に伝送する。これによりディスプレイ装置6の表示画面上には、検索結果（構造化文書としての文書論理構造）の文書集合のリストが表示される。その表示状態の一例を図9に示す。図9に示す例では、検索結果ウィンドウ91内に、5つの文書のリストが表示されている。

【0073】この表示状態で、キーボード4やマウス5を操作することにより、表示されている文書論理構造（構造化文書）のリストから所望の文書論理構造を指示し、更に読み出しを指示すると、この指示情報が入出力制御装置3に伝送される。この指示に応答した文書検索手段12は、文書構造抽出手段14から転送された文書論理構造の集合のデータから、対応する文書論理構造を選び出す。

【0074】ここで、文書割付け手段15が機能するよう指定されていた場合、文書検索手段12は、選び出した文書論理構造を文書割付け手段15に送出する。文書割付け手段15では、伝送されてきた文書論理構造に対して、その文書論理構造（抽出された文書構造）が属していた構造化文書に対応する文書割付けテンプレートに従って文書割付け構造を付与する。この割付け結果は入出力制御装置3を通じてディスプレイ装置6に表示される。この表示状態の一例を図10に示す。図10に示す例では、「文書データベースとは」のリストが選択されたので、文書ウィンドウ101内に、図8に示した文書論理構造に対して文書割付け構造が付与された文書構造の内容が表示されている。

【0075】一方、文書構造表示手段16が機能するよう指定されていた場合は、選び出された文書論理構造は文書構造表示手段16に伝送されることとなる。すると文書構造表示手段16は、伝送されてきた文書論理構造を所定の表示方法で表示可能な構造にする。その表示可能な構造は入出力制御装置3を通じてディスプレイ装置6に表示される。この表示状態の一例を図11に示す。この図11に示す例においては、表示ウィンドウ111内に、図8に示す文書論理構造における「タイトル」、「著者名」、「段落」の各名前に対応して該当する文書内容を表示するという表示方法に従って、検索結果が表示されている。

【0076】これにより、ディスプレイ装置6の表示画面には検索結果である文書が表示されることとなる。

【0077】上述した様に第1の実施例によれば、構造化文書の検索条件と構造抽出とを指示することにより、複数の構造化文書中から、文書集合を指定する情報に適合する文書集合が検索され、更にその文書集合の各要素（検索された各構造化文書）から、構造化文書の抽出に関する情報に基づいた文書論理構造が抽出されるので、複数の構造化文書からそれぞれ、同一の文書論理構造を有する一部分の内容のみを抽出することができる。

【0078】また文書割付け手段によって、文書構造抽出手段により抽出された文書論理構造が属していた構造化文書に対応する文書割付けテンプレートに従って、文書構造抽出手段により抽出された文書論理構造を割付けるようにしたので、割付け済みの文書を得ることができる。

【0079】更に文書構造表示手段によって、文書構造抽出手段により抽出された文書論理構造の内容を表示するようにしたので、抽出された文書論理構造の内容を視覚的に認識することができる。

【0080】次に、本発明の第2の実施例を図12を参照して説明する。

【0081】図12は、本発明に係る文書データベース装置の第2の実施例を機能ブロック図で示したものである。この機能ブロック図は、図1に示した第1の実施例の機能ブロック図の構成において、文書構造表示手段16を削除し、また、テンプレート格納手段121、テンプレート指定手段122、テンプレート編集手段123、複数の文書構造表示手段124-1、124-2、124-3、表示選択手段1250を追加した構成になっている。なお、この図において、図1に示す構成要素の同様の機能を果たす部分には同一の符号を付している。

【0082】テンプレート格納手段121は、例えば磁

気ディスクを備えて構成されるファイリングシステムであって、異なったレイアウト構造を得るために用いられる複数の文書割付テンプレートを格納している。ここには、例えば図16に示したような文書割付テンプレートも格納されている。

【0083】テンプレート指定手段122は、文書割付け手段15が割付けを行うときの文書割付けテンプレートを指定するためのものであり、ユーザによって指定された文書割付けテンプレートを文書割付け手段15に指示する。このとき、ユーザから指定された指示情報に基づいて、テンプレート格納手段121を検索して、実際の文書割付けテンプレートを示す情報を得る。

【0084】テンプレート編集手段123は、テンプレート格納手段121中の文書割付けテンプレートを変更するものであり、ユーザからの編集対象の文書割付けテンプレートの指示情報及び削除／変更の指示情報に基づいて、テンプレート格納手段121中の文書割付けテンプレートに対して削除／変更の操作を実行する。またテンプレート編集手段123は、ユーザからの文書割付けテンプレートの作成指示情報に従って、新たな文書割付けテンプレートを作成し、これをテンプレート格納手段121に格納する。

【0085】複数の文書構造表示手段124-1、124-2、124-3は、割付け構造を持たない構造化文書つまり文書論理構造の表示を行うためのものであり、抽出された文書論理構造をディスプレイ装置6上に表示させる。各文書構造表示手段は、それぞれ異なった表示形式で文書論理構造である文書構造を表示する。

【0086】表示選択手段125は、表示形式と文書構造表示手段との対応関係について予め知っており、ユーザからの表示形式を示す入力情報に応じて、上記各文書構造表示手段中から所望の文書構造表示手段を選択する。なおこの実施例では、複数の表示形式を示す入力情報が入力された場合は、表示選択手段125は、各表示形式に対応する文書構造表示手段を選択することになる。従って、単数又は複数の表示形式で文書論理構造を表示することができる。

【0087】係る構成において、文書データベース装置の文書検索処理について説明する。なおこの第2の実施例においては、検索条件（検索式）に適合する文書の検索処理は、基本的には上記第1の実施例の処理と同様であるので、ここでは、その説明については省略する。

【0088】この第2の実施例が、第1の実施例と異なる点は、大別して次の3点である。

（1）複数の文書割付けテンプレート中から、所望の文書割付けテンプレートの規則に従って文書論理構造を割付ける。

（2）複数の文書構造表示手段中から、所望の文書構造表示手段によって文書論理構造を表示する。

（3）新たな文書割付けテンプレートを追加したり、既存の文書割付けテンプレートの削除／変更を行う。

【0089】次に、これらの処理動作について説明する。

【0090】抽出された文書論理構造に対して割付け処理を施した結果を表示させたい場合は、ユーザは、キーボード4あるいはマウス5を操作して、少なくとも、「検索式」と「文書論理構造に対する割付け処理を行うべく割付け指示情報」と「所望の文書割付けテンプレートの指示情報」とを設定する。なお文書割付けテンプレートの指定がなかった場合は予め設定された文書割付けテンプレートが適用されるようになっている。

【0091】このようなユーザからの入力情報のうち、所望の文書割付けテンプレートの指示情報は入出力制御装置3を経てテンプレート指定手段122入力される。テンプレート指定手段122では、その情報に基づいて、テンプレート格納手段121内から、実際に適用される文書割付けテンプレートを検索し、そのテンプレートを示す情報（例えば、テンプレートを識別する識別情報、テンプレートが格納されているアドレス情報など）を文書割付け手段15に通知する。

【0092】これに対し、ユーザからの入力情報のうち、検索式及び割付け指示情報は、入出力制御装置3を経て検索式解析手段11に入力される。検索式解析手段11では、検索式情報に基づいて上述した第1の実施例と同様の処理を実行し、この結果を文書検索手段12に通知する。割付け指示情報については、そのまま文書検索手段12に通知する。このとき文書検索手段12では、割付け指示情報が入力されたので、検索式に適合した検索結果を文書割付け手段15に通知するということを認識する。なお検索式に適合する検索結果を得るための検索処理は、上述した第1の実施例と同様である。

【0093】ここで、例えば、図16に示したような文書割付けテンプレートが、テンプレート格納手段121に格納され、指定されたとする。また、第1の実施例で説明した様に図2に示した検索式が入力され、結果として文書構造抽出手段14によって図8に示した様な文書論理構造が抽出され、更に文書検索手段による検索処理が終了して、図9に示すような検索結果ウィンドウ91が表示され、その後、ユーザによって検索結果ウィンドウ91内の「文書データベースとは？」が選択されたとする。

【0094】このような前提においては、文書論理構造は文書検索実行手段12に入力されることになるので、文書検索実行手段12では、入力された文書論理構造を文書割付け手段15に通知する。すると文書割付け手段15は、既に通知されている文書割付けテンプレートを示す情報に基づいて、テンプレート格納手段121から、適用すべき文書割付けテンプレート（この例では図16に示したような文書割付けテンプレート）を読み出すとともに、その文書割付けテンプレートに従って、入

力された文書論理構造を割り付ける。この割付け結果は、入出力制御装置３を経てディスプレイ装置６に入力されるので、このディスプレイ装置６には、図１０に示す文書ウィンドウ１０１の様な結果が表示される。

【００９５】次に、割付け構造を持たない構造化文書つまり文書論理構造をそのまま表示させたい場合は、ユーザは、キーボード４或いはマウス５を操作して、少なくとも、「検索式」と「文書論理構造の表示処理を行うべく表示指示情報」と「所望の表示形式を示す情報」とを設定する。なお表示形式の指定がなかった場合は予め設定された表示形式で表示されるようになっている。

【００９６】このようなユーザからの入力情報のうち、所望の表示形式を示す情報は入出力制御装置３を経て表示選択手段１２５に入力される。表示選択手段１２５では、その情報に基づいて、文書構造表示手段１２４-1、１２４-2、１２４-3の中から所望の文書構造表示手段を選択する。

【００９７】これに対し、ユーザからの入力情報のうち、検索式及び表示指示情報は入出力制御装置３を経て検索式解析手段１１に入力される。検索式解析手段１１では、検索式に基づいて上述した第１の実施例と同様の処理を実行し、この結果を文書検索手段１２に通知する。また表示指示情報については、そのまま文書検索手段１２に通知する。このとき文書検索手段１２では、表示指示情報が入力されたので、検索式に適合する検索結果を、文書構造表示手段に通知するということを認識する。なお検索式に適合する検索結果を得るための検索処理は、上述した第１の実施例と同様である。

【００９８】ここで、例えば、文書構造表示手段１２４-1が文書論理構造をテーブル形式で表示するように設定され、文書構造表示手段１２４-2が文書論理構造を木構造形式で表示するように設定され、文書構造表示手段１２４-3が文書論理構造をハイパーテキスト形式で表示するように設定されていたとする。また第１の実施例で説明した様に図２に示した検索式が入力され、結果として文書構造抽出手段１４によって図８に示した様な文書論理構造が抽出され、更に、文書検索手段による検索処理が終了して、図９に示した様な検索結果ウィンドウ９１が表示され、その後、ユーザによって検索結果ウィンドウ９１内の「文書データベースとは？」が選択されたとする。

【００９９】このような前提においては、文書論理構造は文書検索実行手段１２に入力されることになるので、文書検索実行手段１２では、入力された文書論理構造を各文書構造表示手段に通知する。ここで文書構造表示手段１２４-1が選択されていたとすれば、文書構造表示手段１２４-1は、文書論理構造を表示可能な構造にすべくテーブル形式の構造に変更する。このテーブル形式の構造は、入出力制御装置３を通してディスプレイ装置６に表示される。このとき、ディスプレイ装置６には、図１

１に示した表示ウィンドウ１１１が表示される。

【０１００】また文書構造表示手段１２４-2が選択されていた場合は、文書論理構造は、図８に示した様な構造に対応する木構造として表示ウィンドウ内に表示されることになる。また文書構造表示手段１２４-3が選択されていた場合は、文書論理構造は、図８に示した様な構造に対応するハイパーテキストとして表示ウィンドウ内に表示されることになる。さらに複数の文書構造表示手段が選択されていた場合は、それぞれの表示手段の表示形式に対応した検索結果が、それぞれの表示ウィンドウ内に表示される。

【０１０１】最後に、文書割付けテンプレートの編集処理について説明する。

【０１０２】ユーザからの文書割付けテンプレートの作成指示情報、編集対象の文書割付けテンプレートの指示情報及び削除／変更の指示情報等の編集処理操作に関する情報は、入出力制御装置３を経て、テンプレート編集手段１２３に入力される。テンプレート編集手段１２３では、編集対象の文書割付けテンプレートの指示情報及び削除／変更の指示情報に基づいて、テンプレート格納手段１２１中の文書割付けテンプレートに対して削除／変更の操作を実行すると共に、文書割付けテンプレートの作成指示情報に従って、新たな文書割付けテンプレートを作成し、これをテンプレート格納手段１２１に格納する。

【０１０３】上述したように第２の実施例によれば、文書割付け手段が、テンプレート指定手段により指定された文書割付けテンプレートに従って、文書構造抽出手段により抽出された文書論理構造を割付けるようにしたので、割付け済みの文書を得ることができる。

【０１０４】また、文書テンプレート編集手段が既存の文書割付けテンプレートを編集、新たな文書割付けテンプレートを作成するようにし、文書割付け手段が、作成又は編集された文書割付けテンプレートに従って、文書構造抽出手段により抽出された文書論理構造を割付けるようにしたので、所望のレイアウトに応じた割付け済みの文書を得ることができる。

【０１０５】更に、選択表示選択手段によって選択された単数又は複数の文書構造表示手段によって、文書構造抽出手段により抽出された文書論理構造の内容を表示するようにしたので、所望の表示形式で、抽出された文書論理構造の内容を視覚的に認識することができる。

【０１０６】

【発明の効果】以上説明したように本発明によれば、検索対象の構造化文書から、構造化文書の抽出に関する情報に基づいた文書構造が抽出されるので、構造化文書の一部分のみを抽出することができるという利点がある。

【０１０７】また、複数の構造化文書中から、文書集合を指定する情報に適合する文書集合が検索され、更に、その文書集合の各要素（検索された各構造化文書）か

ら、構造化文書の抽出に関する情報に基づいた文書構造が抽出されるので、複数の構造化文書からそれぞれ、同一の文書構造を有する一部分の内容のみを抽出することができるという利点がある。

【０１０８】また、抽出された文書構造の内容を割付けるようにしたので、割付け済みの文書を得ることができるという利点がある。

【０１０９】また、抽出された文書構造の内容を表示するようにしたので、抽出された文書構造の内容を視覚的に認識することができるという利点がある。

【０１１０】また、選択された単数又は複数の文書構造表示手段によって、抽出された文書構造の内容を表示するようにしたので、所望の表示形式で、抽出された文書構造の内容を視覚的に認識することができるという利点がある。

【０１１１】また、抽出された文書構造を、その文書構造が属していた構造化文書に対応する予め設定された文書割付けテンプレートに従って割付けるようにしたので、割付け済みの文書を得ることができるという利点がある。

【０１１２】また、文書割付け手段が、テンプレート指定手段により指定された文書割付けテンプレートに従って、文書構造抽出手段により抽出された文書構造を割付けるようにしたので、割付け済みの文書を得ることができるという利点がある。

【０１１３】更には、抽出された文書構造を、編集された文書割付けテンプレートに従って割付けるようにしたので、所望のレイアウトに応じた割付け済みの文書を得ることができるという利点がある。

【０１１４】以上のことから、文書データベースに蓄積されている構造化文書から一部分の文書内容を抽出することができると共に、該抽出した内容の割付け処理を行うことができ、かつ、伝送データ量、伝送時間、文書編集時間及び文書割付け時間を軽減することのできる文書データベース装置を提供することができるという効果を奏する。

【図面の簡単な説明】

【図１】本発明に係る文書データベース装置の第１の実施例を示す機能ブロック図。

【図２】本実施例において用いられた検索式を示す図。

【図３】図２に示した検索式を解析して得られた解析木を示す図。

【図４】本実施例における構造化文書へのポインタの集合を示す図。

【図５】本実施例における構造化文書のファイル内部へポインタを張るための情報の読み込みを説明するための図。

【図６】本実施例における解析木と構造化文書とがリンク付けされた対応関係を示す図。

【図７】図６に示したリンク付けに基づいてコピーされた文書の論理構造を示す図。

【図８】図２に示した検索式に基づいて抽出された文書の論理構造の一例を示す図。

【図９】本実施例における検索結果の文書集合のリストの一例を示す図。

【図１０】本実施例における抽出された文書の文書論理構造に対して文書の割付け構造を付与した検索結果の表示例を示す図。

【図１１】本実施例における抽出された文書の文書論理構造（文書の割付け構造無し）の検索結果の表示例を示す図。

【図１２】本発明に係る文書データベース装置の第２の実施例を示す機能ブロック図。

【図１３】構造化文書の一例を示す図。

【図１４】図１３に示した構造化文書の文書割付け構造で表現されるレイアウトイメージを示す図。

【図１５】図１３に示した構造化文書の内部表現（文書構造）を示す図。

【図１６】割付けテンプレートの一例を示す図。

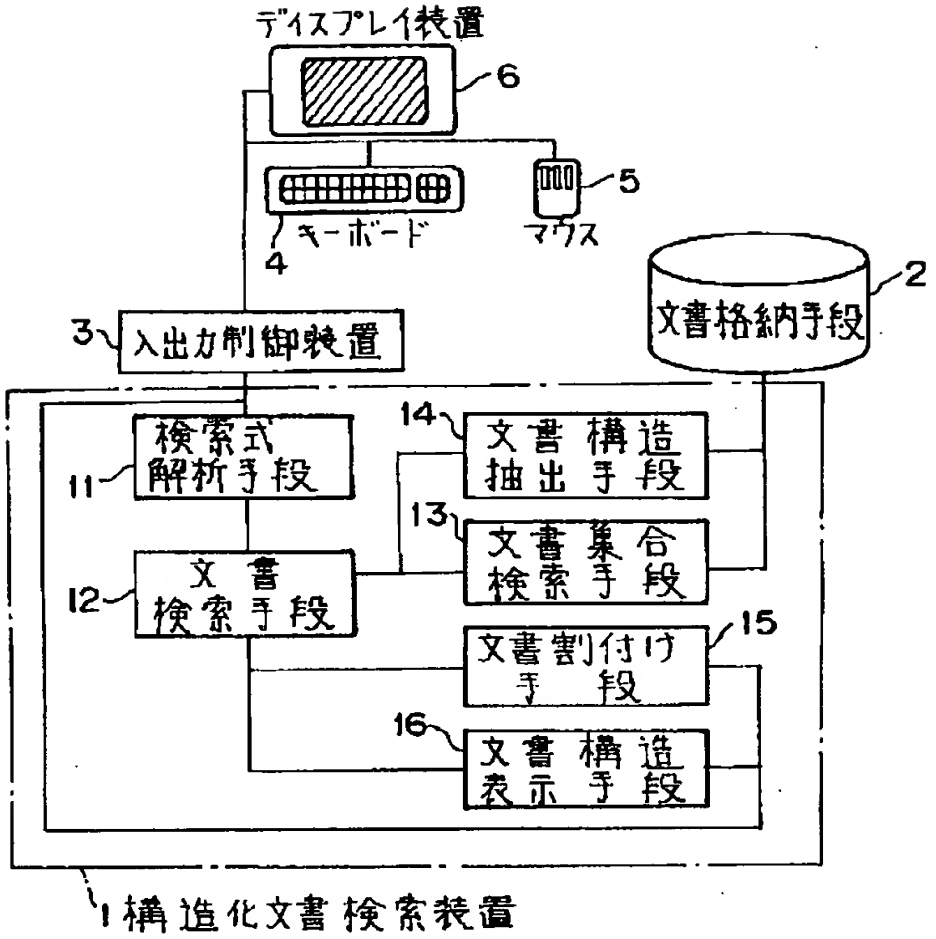
【図１７】図１３に示した構造化文書の文書論理構造を示す図。

【図１８】図１３に示した構造化文書の文書割付け構造を示す図。

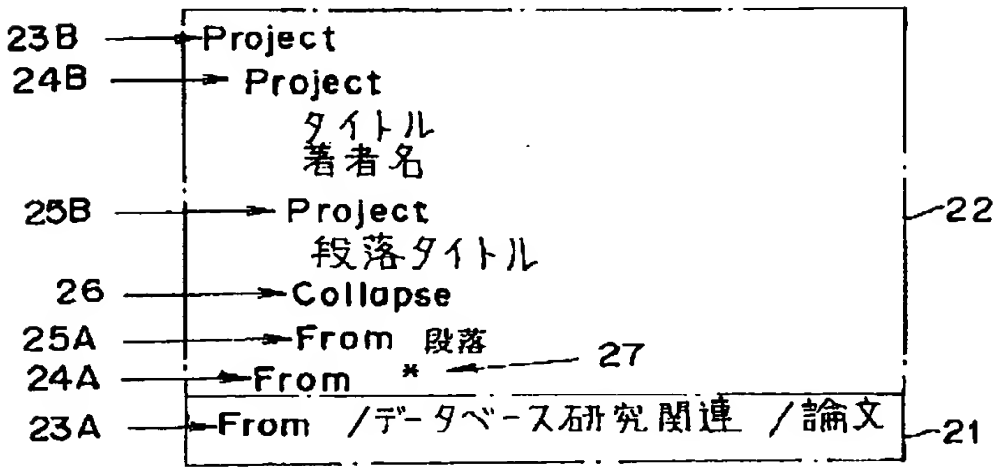
【符号の説明】

１…構造化文書検索装置、２…文書格納手段、３…入出力制御装置、４…キーボード、５…マウス、６…ディスプレイ装置、１１…検索式解析手段、１２…文書検索手段、１３…文書集合検索手段、１４…文書構造抽出手段、１５…文書割付け手段、１６、１２４-1、１２４-2、１２４-3…文書構造表示手段。１２１…テンプレート格納手段、１２２…テンプレート指定手段、１２３…テンプレート編集手段、１２５…表示選択手段。

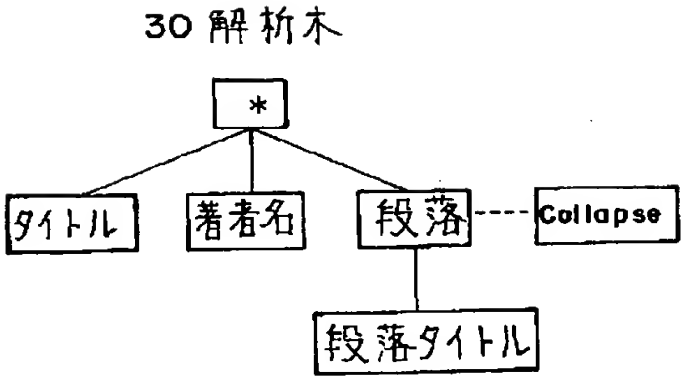
【図1】


ディスプレイ装置
6
5
キーボード
マウス
4
2
文書格納手段
3
入出力制御装置
14
文書構造抽出手段
検索式解析手段
11
13
文書集合検索手段
文書検索手段
12
15
文書割付け手段
16
文書構造表示手段
1 構造化文書検索装置

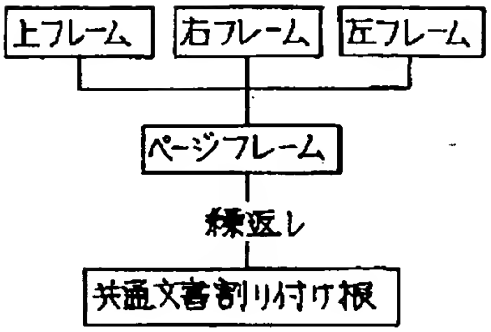
【図2】


23B
Project
24B
Project
タイトル
著者名
25B
Project
段落タイトル
22
26
Collapse
25A
From 段落
24A
From *
27
23A
From /データベース研究関連 /論文
21

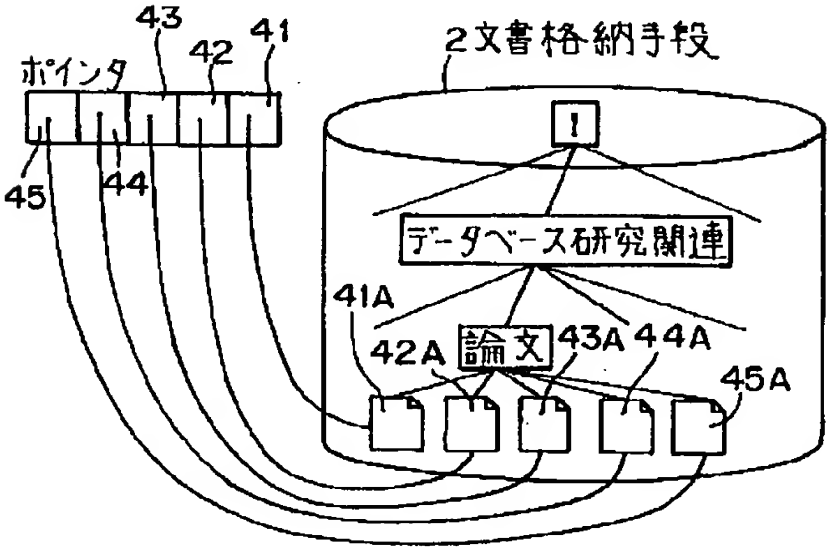
【図３】



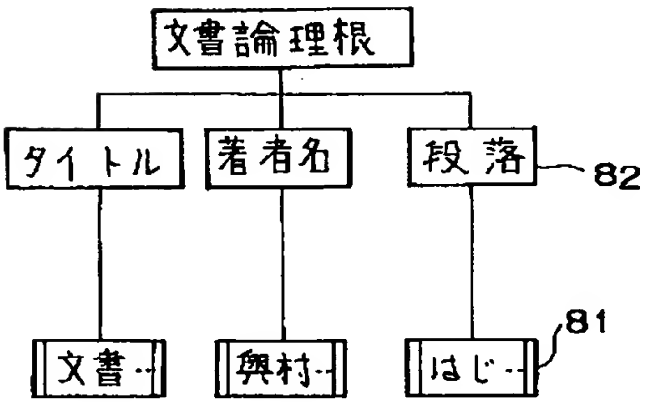
【図１６】



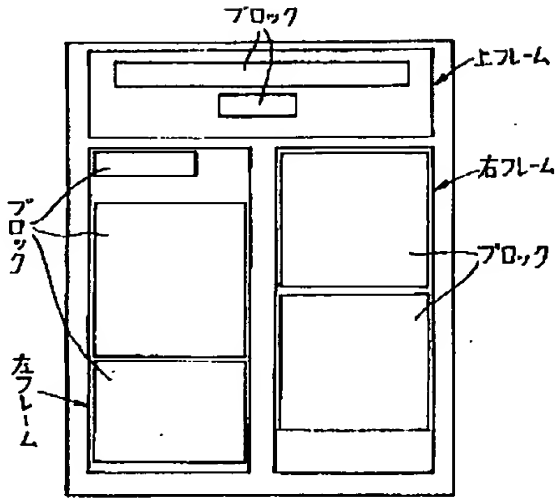
【図４】



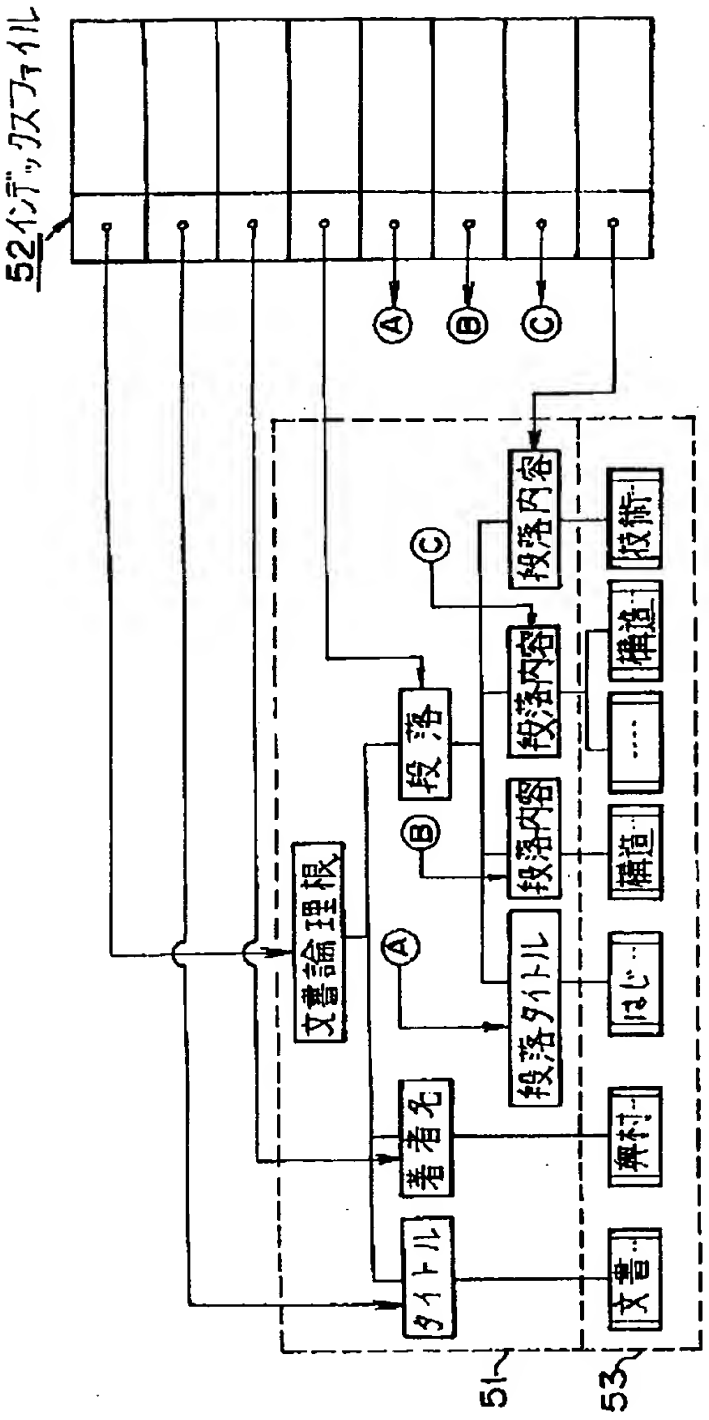
【図８】



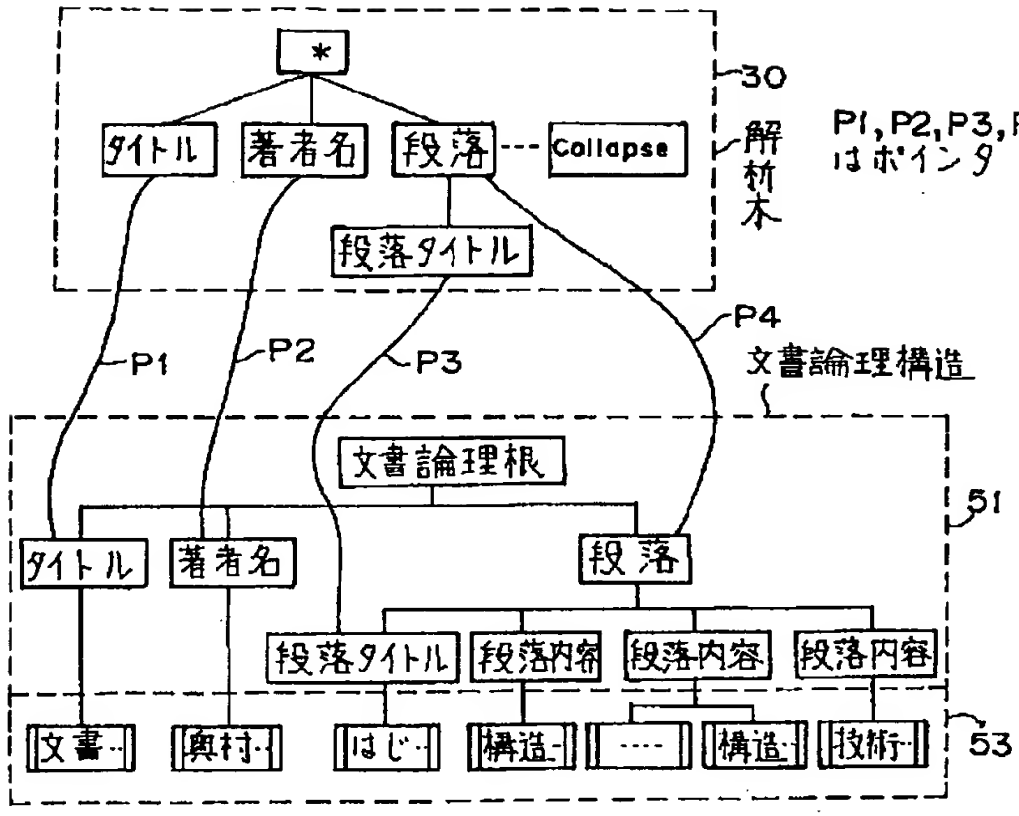
【図１４】



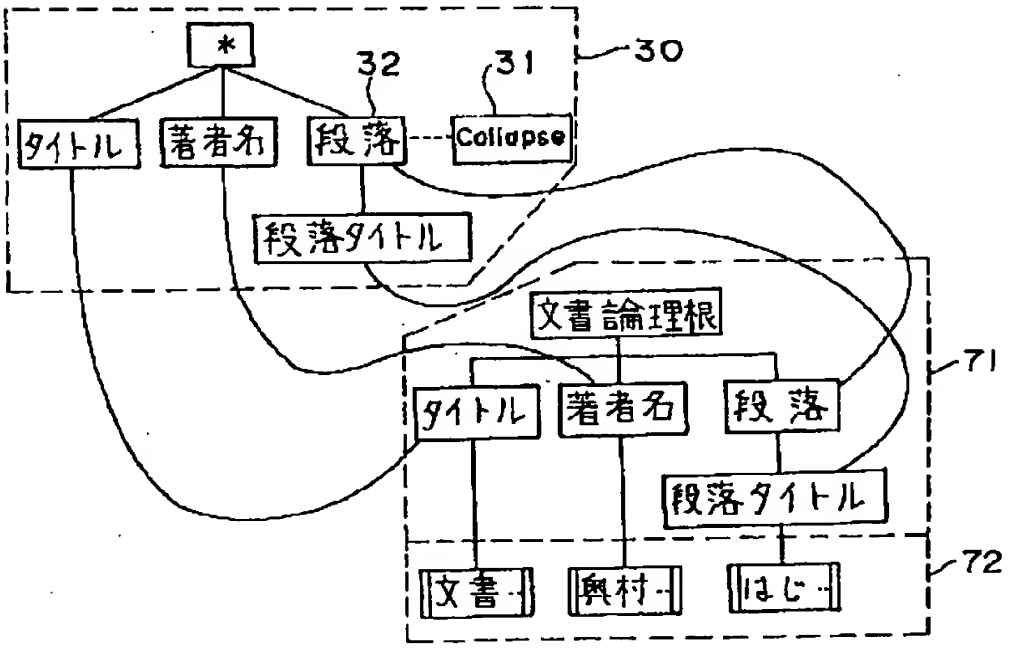
【図５】


52インデックスファイル
Ⓐ
Ⓑ
Ⓒ
文書論理根
タイトル
著者名
段落
段落タイトル
段落内容
段落内容
段落内容
文書…
著者…
はじ…
構造…
……
構造…
技術…
51
53

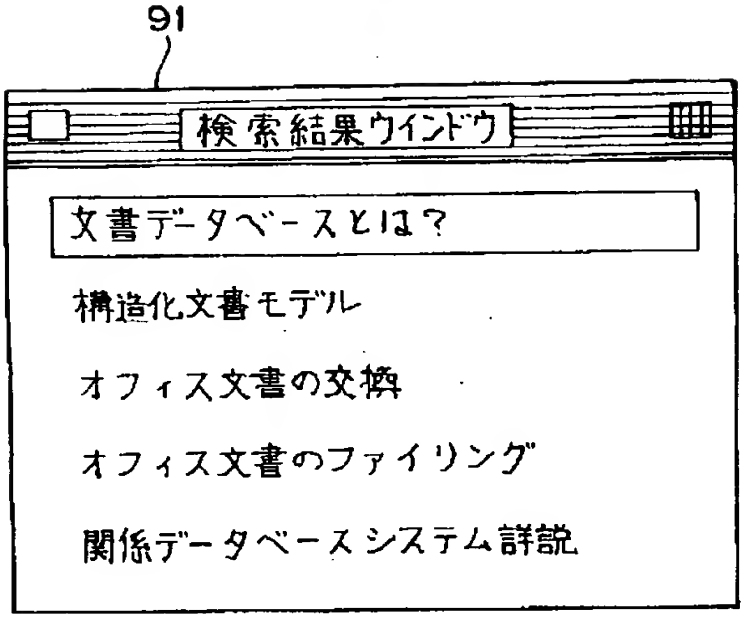
【図6】


*
タイトル
著者名
段落
Collapse
段落タイトル
30
解析木
P1,P2,P3,P4
はポインタ
P1
P2
P3
P4
文書論理構造
文書論理根
タイトル
著者名
段落
段落タイトル
段落内容
段落内容
段落内容
51
文書
奥村
はじ
構造
構造
技術
53

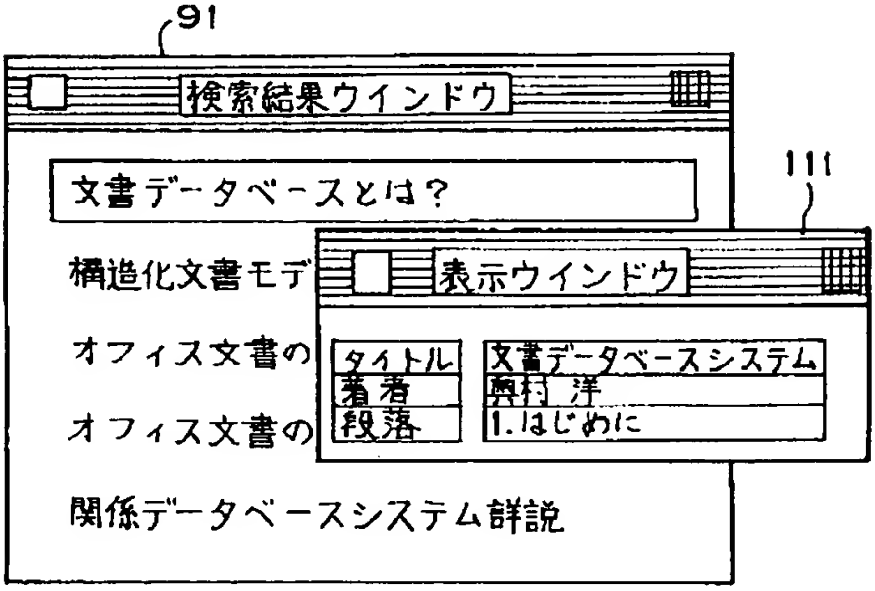
【図7】


*
32
31
30
タイトル
著者名
段落
Collapse
段落タイトル
文書論理根
71
タイトル
著者名
段落
段落タイトル
72
文書
奥村
はじ

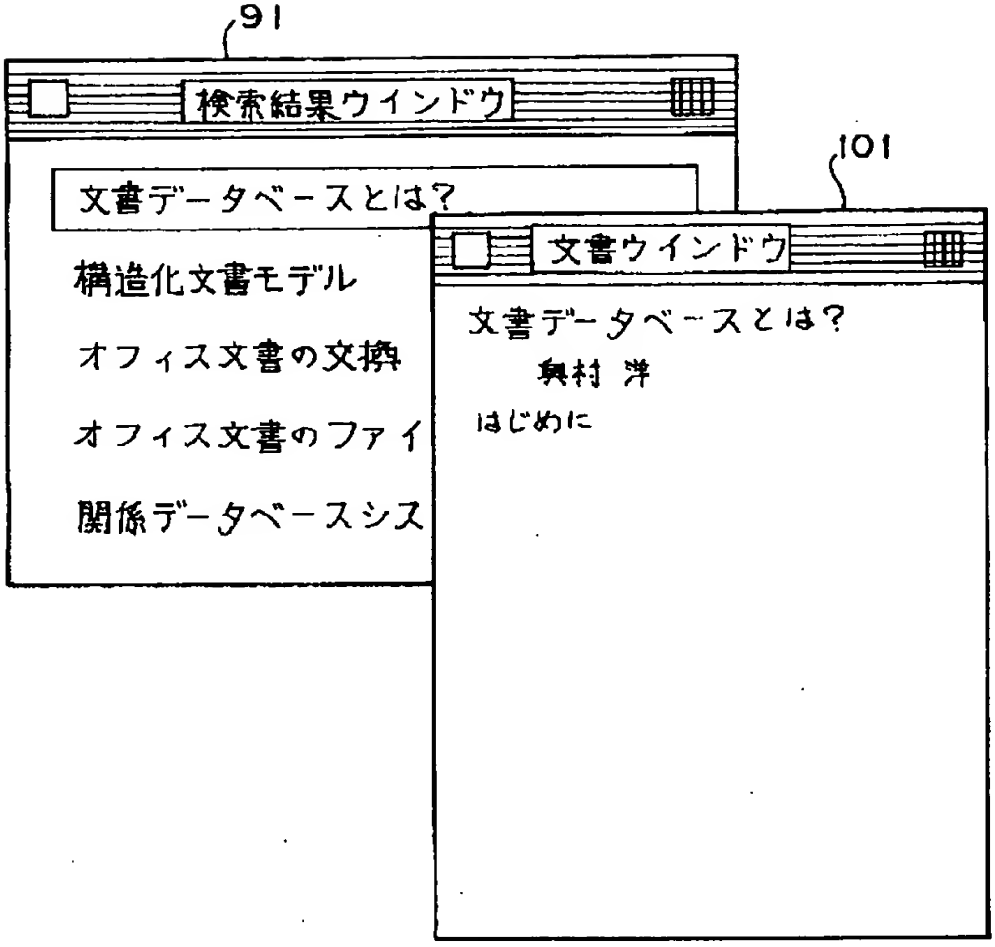
【図9】


91
検索結果ウインドウ
文書データベースとは？
構造化文書モデル
オフィス文書の交換
オフィス文書のファイリング
関係データベースシステム詳説

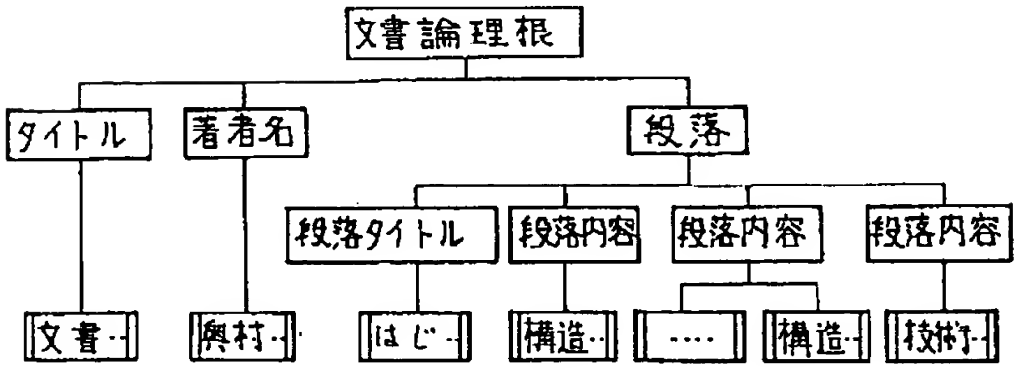
【図11】


91
検索結果ウインドウ
文書データベースとは？
111
構造化文書モデ
表示ウインドウ
オフィス文書の
タイトル
文書データベースシステム
著者
奥村 洋
オフィス文書の
段落
1.はじめに
関係データベースシステム詳説

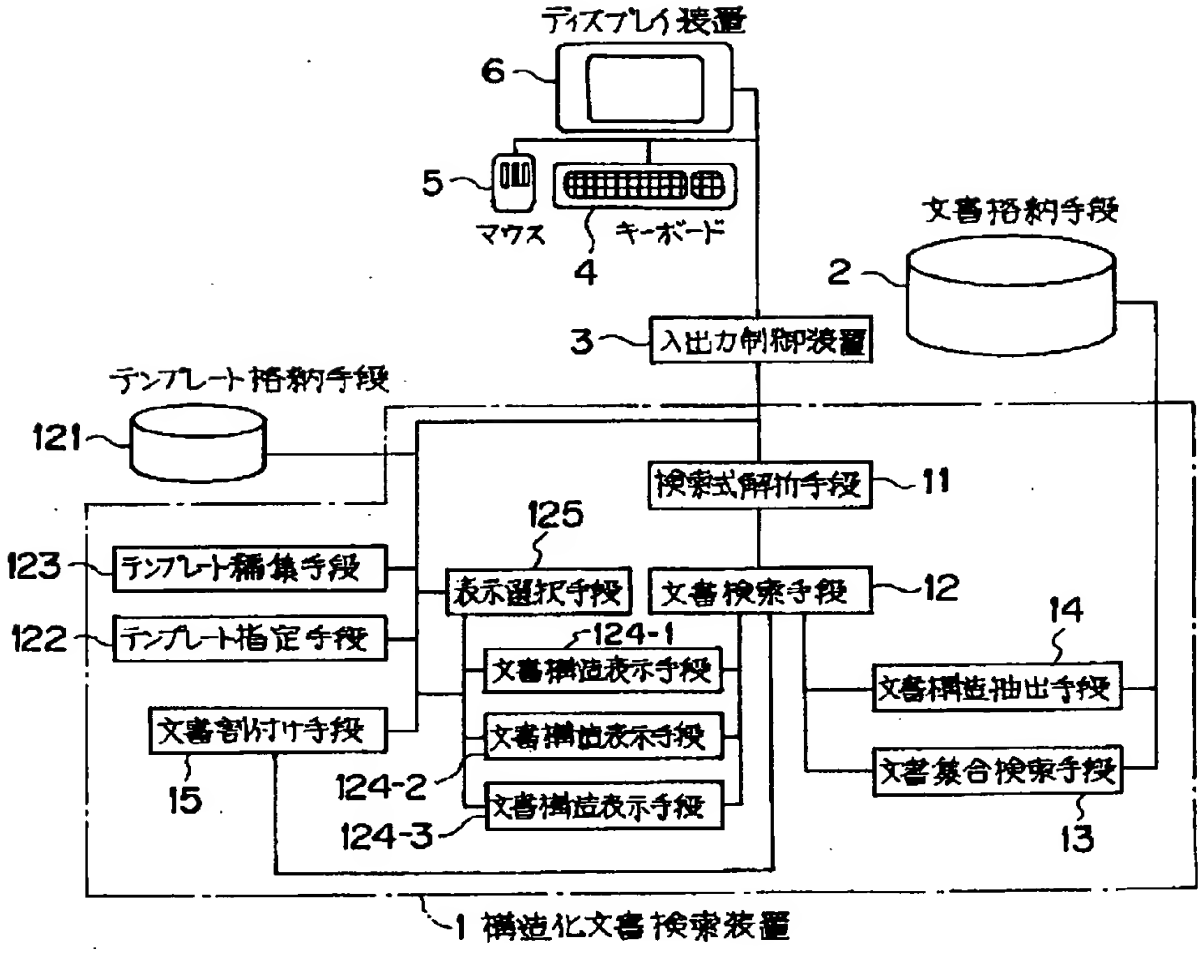
【図１０】


91
検索結果ウインドウ
文書データベースとは?
構造化文書モデル
オフィス文書の交換
オフィス文書のファイ
関係データベースシス
101
文書ウインドウ
文書データベースとは?
奥村 洋
はじめに

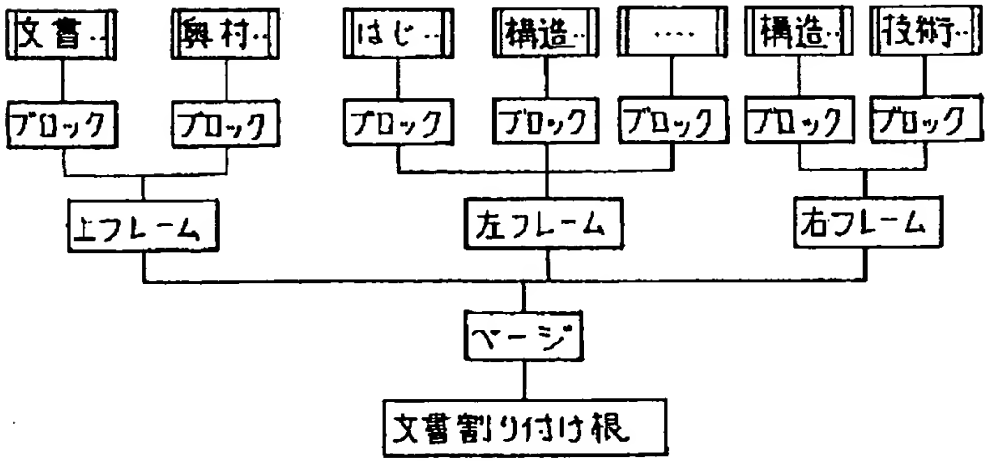
【図１７】


文書論理根
タイトル
著者名
段落
段落タイトル
段落内容
段落内容
段落内容
文書…
奥村…
はじ…
構造…
……
構造…
技術…

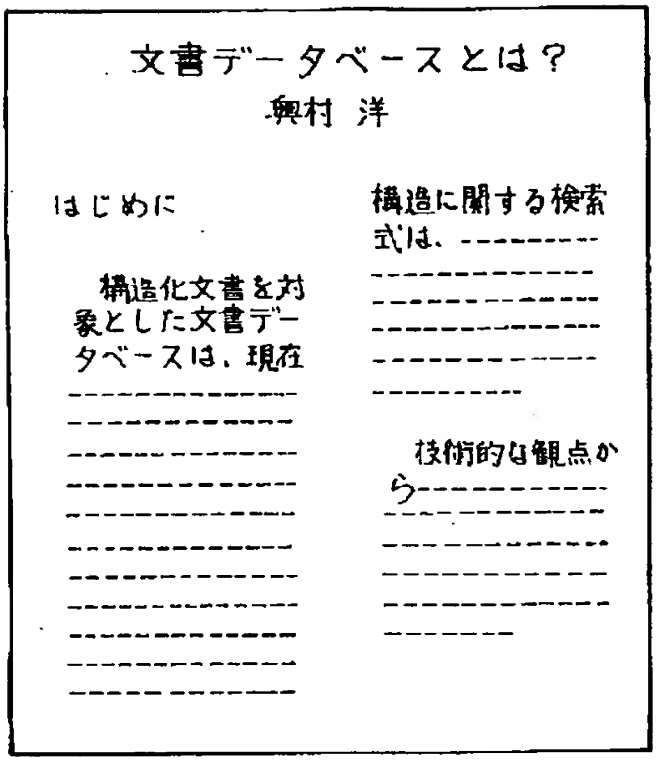
【図１２】


ディスプレイ装置
6
5
マウス
キーボード
4
文書格納手段
2
3
入出力制御装置
テンプレート格納手段
121
検索式解析手段
11
125
123
テンプレート編集手段
表示選択手段
文書検索手段
12
122
テンプレート指定手段
124-1
14
文書構造表示手段
文書構造抽出手段
文書割付け手段
文書構造表示手段
文書集合検索手段
15
124-2
124-3
文書構造表示手段
13
1 構造化文書検索装置

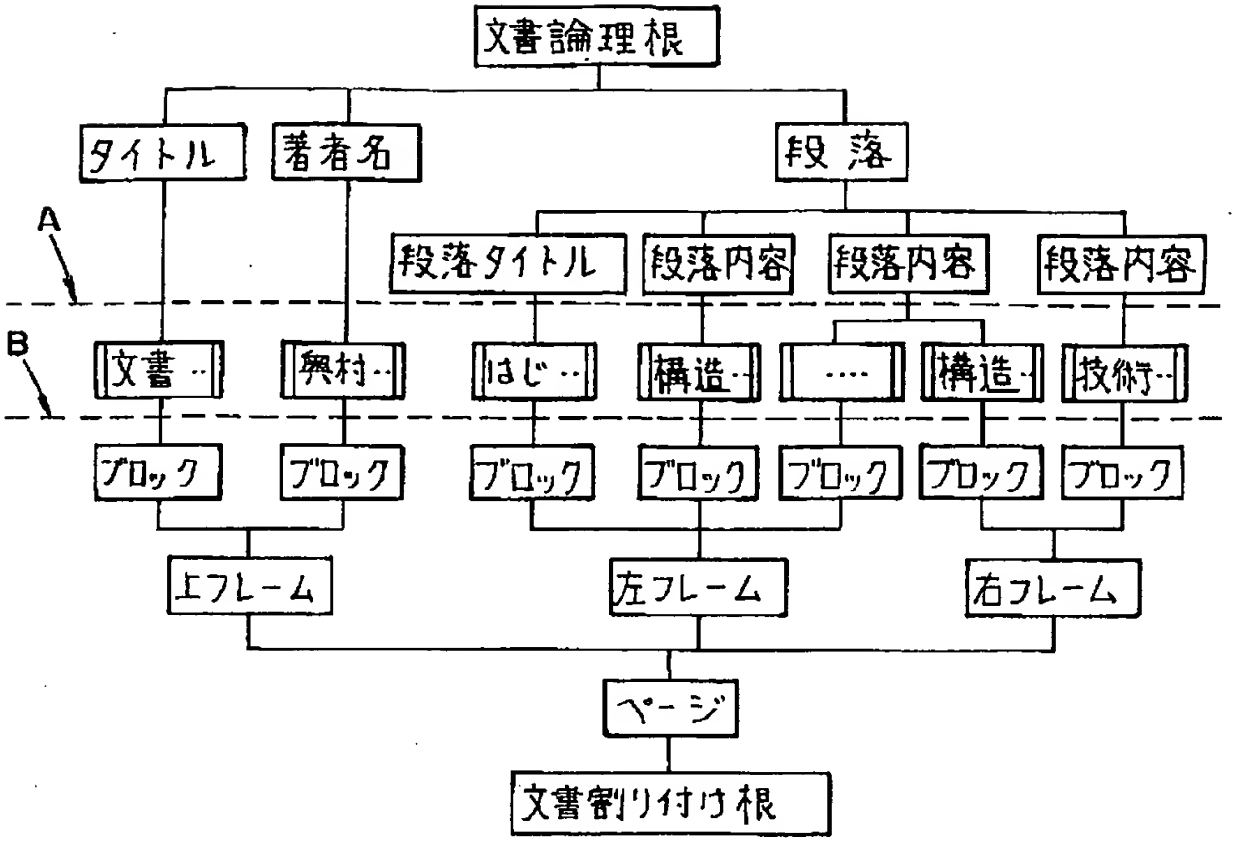
【図１８】


文書..
興村..
はじ..
構造..
....
構造..
技術..
ブロック
ブロック
ブロック
ブロック
ブロック
ブロック
ブロック
上フレーム
左フレーム
右フレーム
ページ
文書割り付け根

【図１３】


文書データベースとは？
興村 洋
はじめに
構造化文書を対象とした文書データベースは、現在
構造に関する検索式は、
技術的な観点から

【図１５】


文書論理根
タイトル
著者名
段落
A
段落タイトル
段落内容
段落内容
段落内容
B
文書‥
興村‥
はじ‥
構造‥
‥‥
構造‥
技術‥
ブロック
ブロック
ブロック
ブロック
ブロック
ブロック
ブロック
上フレーム
左フレーム
右フレーム
ページ
文書割り付け根